Physica A 392 (2013) 1756–1761

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

DCCA cross-correlation coefficient differentiation: Theoretical and practical approaches

G.F. Zebende [a,b,*], M.F. da Silva [a], A. Machado Filho [a]

[a] Computational Modeling Program - SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil
[b] Department of Physics - UEFS 44036-900 Feira de Santana, Bahia, Brazil

ARTICLE INFO

Article history:
Received 30 August 2012
Received in revised form 24 November 2012
Available online 7 January 2013

Keywords:
DFA method
DCCA method
DCCA cross-correlation coefficient

ABSTRACT

We propose in this paper to establish a well-defined relationship between α_{DFA} (the long range auto-correlation exponent) and λ_{DCCA} (the long range cross-correlation exponent), respectively described by the DFA and DCCA methods. This relationship will be accomplished theoretically by differentiating the DCCA cross-correlation coefficient, $\rho_{DCCA}(n)$. Also, for some specific times series, we apply this theory in order to establish its validity.

1. Introduction

In recent decades the study of complex systems has become increasingly important. These systems are studied in many areas of the natural sciences, mathematics, and the social sciences [1–4]. Complex systems have nonlinear behavior, and can be studied by taking into account the properties of fractals [4], such as self-affinity. If, for example, in a given time series $\{y(i)\}$ [5] self-affinity appears, then long range power-law correlations are present [6–10]. This makes the study of complex systems very interesting, because it is possible to identify a universality in different kinds of problems [11,12]. It is known that, in the real world, data are highly nonstationary [13], and many conventional methods of analysis are not suited for nonstationary time series. To address this problem and quantify the long range power-law correlations embedded in a nonstationary time series, the method of detrended fluctuation analysis (DFA) was proposed [14]. The DFA method provides a relationship between $F_{DFA}(n)$ (root mean square fluctuation) and the time scale n, characterized for a power-law:

$$F_{DFA}(n) \propto n^{\alpha}. \tag{1}$$

The scaling exponent α is a self-affinity parameter, representing the long range power-law auto-correlation properties of the time series. If the signal has only short-range correlations (or no correlations at all), $\alpha = 0.5$. On the other hand, if $\alpha < 0.5$, the correlations in the signal are antipersistent, and if $\alpha > 0.5$, the correlations in the signal are persistent.

Furthermore, many observables can be measured and recorded simultaneously, at successive time intervals, forming time series with the same length N [5]. For example, if we have two time series $\{y_1(i)\}$ and $\{y_2(i)\}$, the analysis of the cross-correlation between these time series can be applied. In this case, we can apply a generalization of the DFA method, called the method of detrended cross-correlation analysis (DCCA) [15], to study the long range cross-correlations in the presence of nonstationarity [16]. Thus, given two long range cross-correlated time series, we compute the integrated signals $R_1(k) \equiv \sum_{i=1}^{k} y_1(i)$ and $R_2(k) \equiv \sum_{i=1}^{k} y_2(i)$, where $k = 1, \ldots, N$. Next, we divide the entire time series into $(N - n)$

* Corresponding author at: Computational Modeling Program - SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil. Tel.: +55 157599922788.
E-mail addresses: gfzebende@hotmail.com, gfzebende@pq.cnpq.br (G.F. Zebende).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2013.01.011

overlapping boxes, each containing $(n+1)$ values. For both time series, in each box that starts at i and ends at $i+n$, we define the local trend, $\widetilde{R}_{1,i}(k)$ and $\widetilde{R}_{2,i}(k)(i \le k \le i+n)$, to be the ordinate of a linear least-squares fit. We define the detrended walk as the difference between the original walk and the local trend. Next, we calculate the covariance of the residuals in each box $f^2_{DCCA}(n, i) \equiv 1/(n+1)\sum_{k=i}^{i+n}(R_1(k) - \widetilde{R}_{1,i}(k))(R_2(k) - \widetilde{R}_{2,i}(k))$. Finally, the detrended covariance function is calculated by summing over all overlapping $(N-n)$ boxes of size n,

$$F^2_{DCCA}(n) \equiv (N-n)^{-1}\sum_{i=1}^{N-n} f^2_{DCCA}(n, i). \tag{2}$$

If self-affinity appears, then a power-law exists in the cross-correlations, in other words,

$$F^2_{DCCA}(n) \sim n^{2\lambda}, \tag{3}$$

where λ is the long range power-law cross-correlation exponent. Supposing $(R_1(k) = R_2(k))$, the detrended covariance $F^2_{DCCA}(n)$ reduces to the detrended variance $F^2_{DFA}(n)$ used in the DFA method. According to Podobnik and Stanley [15], in general, λ tends to be the mean of DFA exponents, e.g.,

$$\lambda \approx \frac{(\alpha_1 + \alpha_2)}{2}. \tag{4}$$

The exponent λ quantifies the long range power-law correlations and also identifies seasonality [17]. But, λ does not quantify the level of cross-correlations [22], and there is no clear relationship between λ and α [18–21].

We will show in this paper that it is possible to derive a clear relationship between α and λ if we employ the DCCA cross-correlation coefficient $\rho_{DCCA}(n)$ [22]. For this purpose, the remainder of the present paper has the following structure: In Section 2, there is a brief discussion about the DCCA cross-correlation and its theoretical implementation; Section 3 presents our data, simulations, and results; we close in Section 4 with our conclusions.

2. Theoretical implementation

Below, we introduce the DCCA cross-correlation coefficient $\rho_{DCCA}(n)$ [22]. The cross-correlation coefficient was created in order to quantify the level of cross-correlation between nonstationary time series. The DCCA cross-correlation coefficient is defined as the ratio between the detrended covariance function F^2_{DCCA} and the detrended variance function F_{DFA} of $\{y_1(i)\}$ and $\{y_2(i)\}$, i.e.,

$$\rho_{DCCA}(n) \equiv \frac{F^2_{DCCA}(n)}{F_{DFA1}(n)\, F_{DFA2}(n)}. \tag{5}$$

In this way, Eq. (5) leads us to a new scale of cross-correlation in nonstationary time series. The value of $\rho_{DCCA}(n)$ ranges between $-1 \le \rho_{DCCA}(n) \le 1$ [22,23]. A value of $\rho_{DCCA}(n) = 0$ means there is no cross-correlation, and it splits the level of cross-correlation between the positive and the negative cases (see Table 1 in Ref. [22]). The detrended coefficient $\rho_{DCCA}(n)$ has been tested on simulated and real time series [22,24]. In order to test whether $\rho_{DCCA}(n)$ can be considered statistically significant or not, a statistical test was proposed by [23]. This test was based in the null hypothesis, i.e., assuming that the time series are independent and identically distributed random variables (i.i.d.), with $\alpha_1 = \alpha_2 = 0.5$. As a result, with 95% of confidence level, this test determines the range within which the cross-correlations can be considered statistically significant.

Now, based on the assumption that there is no clear relation between λ and α, and with the particularity that the detrended cross-correlation coefficient $\rho_{DCCA}(n)$ is a dimensionless coefficient, then we will now propose to derive a well-defined relationship between α and λ. This relationship will be applied to two nonstationary time series with long range power-law auto-correlations and with long range power-law cross-correlations,

$$F_{DFA1}(n) = K_1 n^{\alpha_1}, \qquad F_{DFA2}(n) = K_2 n^{\alpha_2}, \tag{6}$$

and

$$F^2_{DCCA}(n) = K_3 n^{2\lambda}. \tag{7}$$

K_1, K_2, and K_3 are nonzero constants. Thus, from Eqs. (5)–(7), we have

$$\rho_{DCCA}(n) = K n^{2\lambda - \alpha_1 - \alpha_2}, \tag{8}$$

with $K \equiv \frac{K_3}{K_1 K_2}$.

Setting $y \equiv \log_{10}\rho_{DCCA}(n)$ and $x \equiv \log_{10}(n)$, then from Eq. (8),

$$\frac{dy}{dx} = 2\lambda - \alpha_1 - \alpha_2. \tag{9}$$

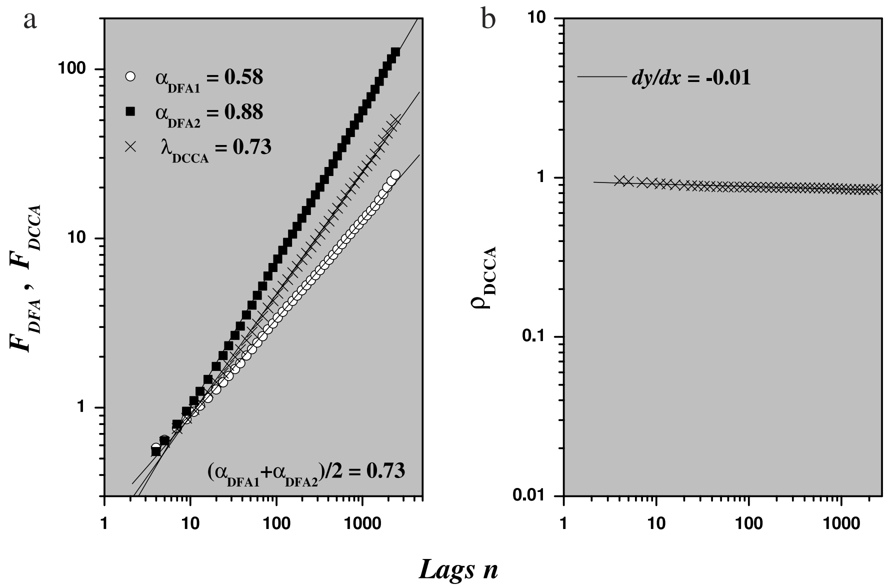

Fig. 1. (a) Detrended variance, F_{DFA}, and detrended covariance, F_{DCCA}, as functions of n (time lag), for two time series generated by ARFIMA processes with $\rho_1 = 0.1$ (○), $\rho_2 = 0.4$ (■), and $W = 1.0$. In this ARFIMA processes both time series, $\{y_1\}$ and $\{y_2\}$, share the same i.i.d. Gaussian process $\varepsilon_{1,i} = \varepsilon_{2,i}$. A long range power-law auto-correlation, $F_{DFA}(n) \sim n^{\alpha}$, and long range power-law cross-correlations, $F_{DCCA}(n) \sim n^{\lambda}$, are generated. (b) DCCA cross-correlation coefficient $\rho_{DCCA} \times n$. Continuous line represents the linear adjust, dy/dx, with $R = 0.99$, $sd = 0.02$, and $p_{value} < 0.0001$.

Therefore, the goal of this paper is establish, using Eq. (9), a clear relationship between α_1, α_2, and λ, via differentiation of ρ_{DCCA}. Looking at Eq. (9), we can see three possibilities [25]:

$$\lambda = (\alpha_1 + \alpha_2)/2,$$
$$\lambda < (\alpha_1 + \alpha_2)/2, \tag{10}$$

and

$$\lambda > (\alpha_1 + \alpha_2)/2.$$

Looking at Eq. (8), we can see that, in the limit of $n \to \infty$, then $1/\log_{10}(n) \to 0$ (very slow decay). Consequently $\lambda \approxeq (a_1 + a_2)/2$, because $(\log_{10} \rho_{DCCA} - \log_{10} K)/\log_{10}(n) \to 0$.

In order to verify these possibilities, we consider here two important cases:

(a) $\frac{dy}{dx} = 0$, with $\lambda = (\alpha_1 + \alpha_2)/2$, which is the typical case;
(b) $\frac{dy}{dx} = \xi$ ($\xi \in \mathbb{R}$). In this case, $\lambda = (\xi + \alpha_1 + \alpha_2)/2$, or $\lambda \neq (\alpha_1 + \alpha_2)/2$.

To test this assumption, listed in items (a) and (b), we propose to analyze some nonstationary time series. This analysis will be done on both simulated and real-world time series.

3. Data and results

We start by treating two times series with long range power-law cross-correlations, using the ARFIMA process [26,27,29]. These time series are independent and uncorrelated, with DFA exponents $\alpha_1 = \alpha_2 = 0.5$ [30]. In the ARFIMA process each variable depends not only on its own past, but also on the past values of the other variable, e.g.,

$$y_{1,i} = W \sum_{n=1}^{\infty} a_n(\rho_1) y_{1,i-n} + (1 - W) \sum_{n=1}^{\infty} a_n(\rho_2) y_{2,i-n} + \varepsilon_{1,i},$$
$$y_{2,i} = (1 - W) \sum_{n=1}^{\infty} a_n(\rho_1) y_{1,i-n} + W \sum_{n=1}^{\infty} a_n(\rho_2) y_{2,i-n} + \varepsilon_{2,i}, \tag{11}$$

and from this process we generate two new time series. In this process $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ denote two independent and identically distributed (i.i.d.) Gaussian variables with zero mean and unit variance, $a_n(\rho)$ are statistical weights defined by $a_n(\rho) = \Gamma(n - \rho)/(\Gamma(-\rho)\Gamma(1 + n))$, where Γ denotes the Gamma function. ρ are parameters ranging from -0.5 to 0.5, related to the DFA exponent, $\alpha = 0.5 + \rho$, and W is a free parameter ranging from 0.5 to 1.0 and controlling the strength of the power-law cross-correlations between $\{y_1(i)\}$ and $\{y_2(i)\}$.

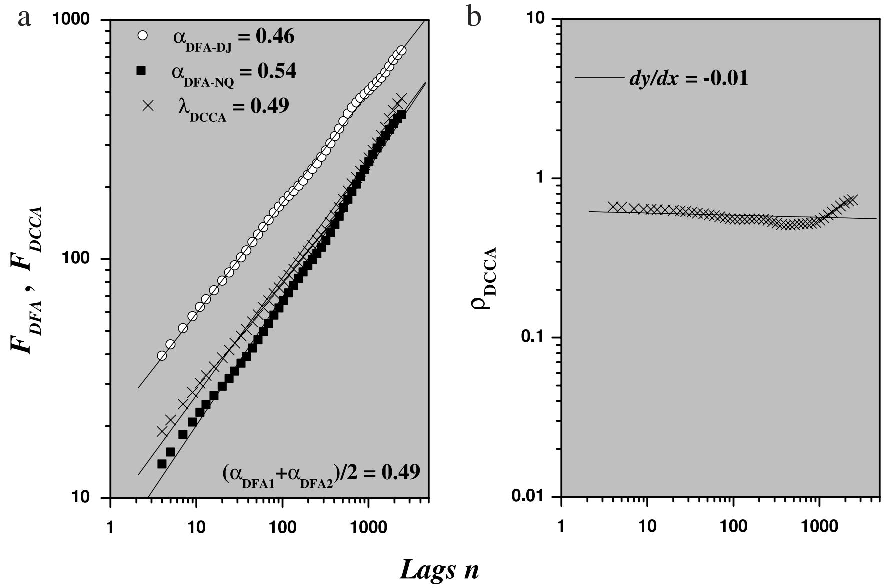

Fig. 2. (a) Detrended variance $F_{DFA}(n)$ and detrended covariance, $F_{DCCA}(n)$ as functions of n (time lag) for Dow Jones $\{y_1\}$ and NASDAQ $\{y_2\}$ differences in closing prices adjusted for dividends and splits, recorded daily between February 5, 1971 and December 9, 2010. (b) This figure represent the log $\times$ log plot of $\rho_{DCCA} \times n$. The continuous line represents the linear adjust, dy/dx, with $R = 0.99$, $sd = 0.02$, and $p_{value} < 0.0001$.

As our first test, we propose to analyze the ARFIMA process where both time series share the same i.i.d. Gaussian process $\varepsilon_{1,i} = \varepsilon_{2,i}$ with $W = 1.0$, see Ref. [28] Appendix A. Podobnik et al. find that these time series are long range power-law cross-correlated, where the scaling cross-correlations exponent λ is equal to the average of the α exponents. These results are also found numerically in Ref. [15]. Using this two-component ARFIMA process with $\rho_1 = 0.1$, $\rho_2 = 0.4$, and $\varepsilon_{1,i} = \varepsilon_{2,i}$, long range cross-correlations are generated (see Fig. 1). As a result of this process, Fig. 1(a), presents the detrended variance, $F_{DFA}(n)$, and the detrended covariance, $F_{DCCA}(n)$, as functions of n. This process can be fitted by power laws, $F_{DFA}(n) \propto n^{\alpha}$ and $F_{DCCA}(n) \propto n^{\lambda}$, with scaling exponents $\alpha_1 = 0.58$, $\alpha_2 = 0.88$, and $\lambda = 0.73$. In this figure we can see that $\lambda = \frac{\alpha_1+\alpha_2}{2}$ (the average of the α exponents). Fig. 1(b) shows the detrended cross-correlation coefficient $\rho_{DCCA}(n)$ in log $\times$ log plot to present the application of Eq. (9). We can see that $\frac{dy}{dx} \simeq 0$ (continuous line), and this illustrates the expected relationship between the exponents: $\lambda = \frac{\alpha_1+\alpha_2}{2}$. To test a real-world case of cross-correlation, with $\frac{dy}{dx} = 0$ and $\lambda = \frac{\alpha_1+\alpha_2}{2}$, we introduce now an empirical application. We choose the time series of some stock markets indexes, because these have been extensively studied [31–34]. Our choice was the study the cross-correlation between the Dow Jones and the NASDAQ, for differences in adjusted closing prices. The data was collected daily over the period between February 5, 1971 and December 9, 2010 [35]. As expected, we find power-law auto-correlations, $F_{DFA}(n) \propto n^{\alpha}$, with $\alpha_{DJ} = 0.46$ (○) and $\alpha_{NQ} = 0.54$ (■), as well as power-law cross-correlations, $F_{DCCA}(n) \propto n^{\lambda}$, with $\lambda = 0.49$ (×) (see Fig. 2(a)). As we can see, $\langle\alpha\rangle = 0.49$ has the same value of λ. Fig. 2(b) illustrates the expected relationship between the exponents, because $\frac{dy}{dx} \approx 0$.

Now to test the case where $\lambda \neq (\alpha_1 + \alpha_2)/2$, an ARFIMA process was generated with $W = 0.85$, $\rho_1 = 0.1$ (time series 1), $\rho_2 = 0.4$ (time series 2), and $\varepsilon_{1,i} \neq \varepsilon_{2,i}$. The results for this simulated case are in Fig. 3.

We can see, Fig. 3(a), that $F_{DFA}(n) \propto n^{\alpha}$ with $\alpha_1 = 0.62$ (○) and $\alpha_2 = 0.72$ (■), for time series 1 and 2. Also, we can see that $F_{DCCA}(n) \propto n^{\lambda}$, with $\lambda = 0.86$ (×). But, the cross-correlation is not perfect for this ARFIMA process [22] (see Fig. 3(b)), and $\lambda \neq (\alpha_1 + \alpha_2)/2$. However, as we can see, $\frac{dy}{dx} = 0.38$, and evidently this simulation confirms the theory suggested in Eq. (9), just because $\lambda = (0.38 + \alpha_1 + \alpha_2)/2$.

Finally, as an empirical example of $\lambda \neq (\alpha_1 + \alpha_2)/2$, we propose to study the cross-correlation between the Dow Jones and the 00062.SS (SSE Shanghai local company) indexes. In order to accomplish this goal, we will study the values of the successive differences in adjusted closing prices from January 4, 2000 to December 9, 2010, recorded daily [35]. Again, as in the previous cases, $F_{DFA}(n)$ and $F_{DCCA}(n)$ can be fitted by power-law correlations with $\alpha_{DJ} = 0.47$(○), $\alpha_{SSE} = 0.59$(■), and $\lambda_{DJ\times SSE} = 0.77$(×), see Fig. 4(a). Fig. 4(b) presents the value of the detrended cross-correlation coefficient, $\rho_{DCCA}(n)$, with $\frac{dy}{dx} = 0.48$ (continuous line). As expected theoretically from Eq. (9), $\lambda = (0.48 + \alpha_{DJ} + \alpha_{SSE})/2$.

4. Conclusions

In this paper, we proposed a new theoretical approach based on the detrended cross-correlation coefficient $\rho_{DCCA}(n)$. The goal in this theory was to use it to establish a relationship between α_{DFA} (the long range auto-correlation exponent) and λ_{DCCA} (the long range cross-correlation exponent). We emphasize that this theory has been tested in both simulated

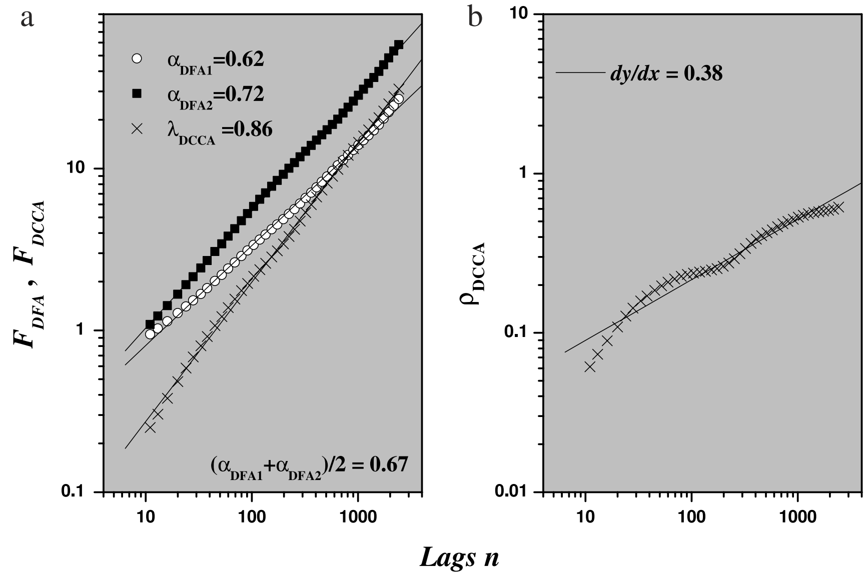

Fig. 3. (a) Detrended variance, $F_{DFA}(n)$, and the detrended covariance, $F_{DCCA}(n)$, as functions of n for two time series $\{y_1\}$ and $\{y_2\}$ generated by the ARFIMA processes: in this process we have $\rho_1 = 0.1$, $\rho_2 = 0.4$, $\varepsilon_i \neq \varepsilon_i'$, and $W = 0.85$. A long range power-law auto-correlation, $F_{DFA}(n) \sim n^{\alpha}$, and long range power-law cross-correlation, $F_{DCCA}(n) \sim n^{\lambda}$, are generated. (b) DCCA cross-correlation coefficient $\rho_{DCCA} \times n$. The continuous line represents the linear adjust, dy/dx, with $R = 0.98$, $sd = 0.05$, and $p_{value} < 0.0001$.

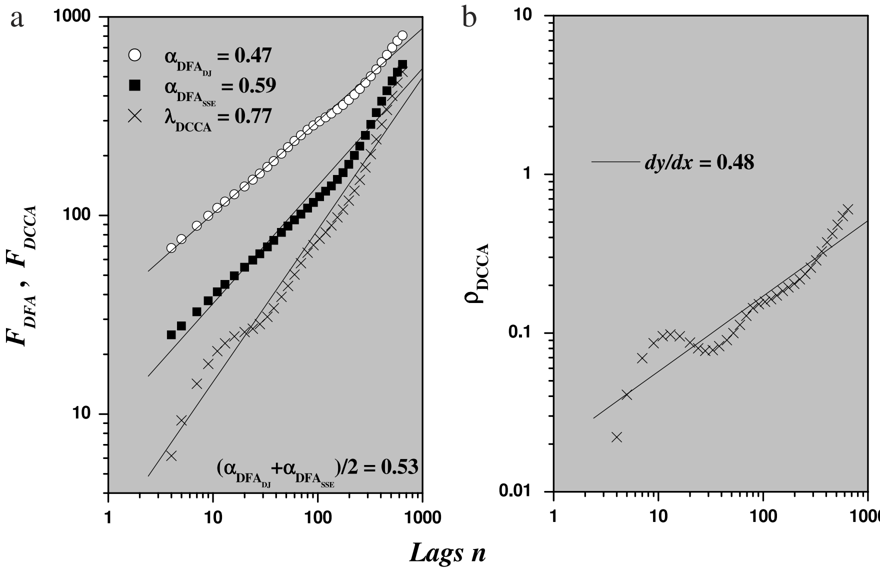

Fig. 4. (a) Detrended variance $F_{DFA}(n)$ and detrended covariance, $F_{DCCA}(n)$ as functions of n for successive differences in adjusted closing prices for Dow Jones y_1 and SSE y_2 indexes, recorded daily from January 4, 2000 to December 9, 2010. (b) DCCA cross-correlation coefficient $\rho_{DCCA} \times n$. The continuous line represents the linear adjust, dy/dx, with $R = 0.94$, $sd = 0.11$, and $p_{value} < 0.0001$.

and real-world cases, and the results are entirely in agreement with Eq. (9). Lastly, this theory can help understand a little better cross-correlations in the presence of nonstationarity, because in the future, the derivative, $\frac{dy}{dx}$, can be an interesting indicator for this.

Acknowledgment

This research received financial support from FAPESB and CNPq (Bahia and Brazilian Research Agencies).

References

[1] Science 284 (5411) (1999) 1–212.

[2] Science 325 (5939) (2009) 357–504.
[3] Nature 410 (6825) (2001) 242–284.
[4] B.B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman and Company, New York, 1982.
[5] http://www.physionet.org/physiobank/database/;
http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html;
http://robjhyndman.com/TSDL/.
[6] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature 356 (1992) 168.
[7] http://www.physionet.org/publications/#papers-with-data.
[8] G.F. Zebende, P.M.C. de Oliveira, T.J.P. Penna, Phys. Rev. E 57 (1998) 3311.
[9] H.E. Hurst, Trans. Amer. Soc. Civ. Eng. 116 (1951) 770.
[10] L.-Y. He, W.-B. Qian, Physica A 391 (2012) 3770.
[11] J. Feder, Fractals, Plenum Press, New York, 1988.
[12] P. Bak, How Nature Works: The Science of Self-Organized Criticality, Oxford University Press, Oxford, 1997.
[13] S.J. Devlin, R. Gnanadesikan, J.R. Kettenring, Biometrika 62 (3) (1975) 531.
[14] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.
[15] B. Podobnik, H.E. Stanley, Phys. Rev. Lett. 100 (2008) 084102.
[16] D. Horvatic, H.E. Stanley, B. Podobnik, Eur. Phys. Lett. 94 (2011) 18007.
[17] G.F. Zebende, A. Machado Filho, Physica A 388 (2009) 4863.
[18] W.-X. Zhou, Phys. Rev. E 77 (2008) 066211.
[19] G. Cao, L. Xu, J. Cao, Physica A 391 (20) (2012) 4855.
[20] L.-Y. He, -P. Chen, Physica A 390 (2011) 297.
[21] L.-Y. He, S.-P. Chen, Physica A 390 (2011) 3806.
[22] G.F. Zebende, Physica A 390 (4) (2011) 614.
[23] B. Podobnik, Z.-Q. Jiang, W.-X. Zhou, H.E. Stanley, Phys. Rev. E 84 (2011) 066118.
[24] R.T. Vassoler, G.F. Zebende, Physica A 391 (7) (2012) 2438.
[25] When $F^2_{DCCA}(n)$ is negative for every n, the constant K_3 is negative. In this case, we can present $-F^2_{DCCA}(n)$ (or simply take K_3 positive) in a log–log plot, in order to calculate the value of λ. In the case where no power-law cross-correlations are present, $\rho_{DCCA}(n) = 0$, or $F^2_{DCCA}(n)$ fluctuates around zero, there does not exist any relation between the exponents α and λ.
[26] J. Hosking, Biometrika 68 (1981) 165.
[27] C.W.J. Granger, J. Econometrics 14 (1980) 227.
[28] B. Podobnik, I. Grosse, D. Horvatic, S. Ilic, P.C. Ivanov, H.E. Stanley, Eur. Phys. J. B 71 (2009) 243.
[29] B. Podobnik, P.C. Ivanov, K. Biljakovic, D. Horvatic, H.E. Stanley, I. Grosse, Phys. Rev. E 72 (2005) 026121.
[30] We built two simulated self-affine times series, with random numbers ($rand_i$) between 0 and 1, $\{y_{rand1}\}$ and $\{y_{rand2}\}$, with $N = 20000$ points. We introduced a cutoff length $l = 10^4$ in our simulations, and we let the sum run from 1 to l; i.e., we set $a_n = 0$ for $n \geq l$.
[31] I. Gvozdanovic, B. Podobnik, D. Wangc, H.E. Stanley, Physica A 391 (2012) 2860.
[32] C.B. Erb, C.R. Harvey, T.E. Viscanta, Financ. Anal. J. 50 (1994) 32.
[33] B. Podobnik, D. Fu, H.E. Stanley, P.C. Ivanov, Eur. Phys. J. B 56 (2007) 47.
[34] B. Podobnik, D. Horvati, A.M. Petersen, H.E. Stanley, Proc. Natl. Acad. Sci. USA 106 (2009) 22079.
[35] http://finance.yahoo.com/.